

Journal of Educational Supervision

Volume 2 | Issue 1

Article 4

2019

Principals' Perceptions of Teacher Evaluation Reform from Structural and Human Resource Perspectives

John Wilson Campbell

Alcoa City School District, jcampbell@alcoaschools.net

Mary Lynne Derrington

University of Tennessee, mderring@utk.edu

Follow this and additional works at: <https://digitalcommons.library.umaine.edu/jes>

Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Educational Leadership Commons](#), [Elementary and Middle and Secondary Education Administration Commons](#), [Elementary Education and Teaching Commons](#), [Junior High, Intermediate, Middle School Education and Teaching Commons](#), and the [Secondary Education and Teaching Commons](#)

Recommended Citation

Campbell, J. W., & Derrington, M. (2019). Principals' Perceptions of Teacher Evaluation Reform from Structural and Human Resource Perspectives. *Journal of Educational Supervision*, 2 (1). <https://doi.org/10.31045/jes.2.1.4>

This Empirical Research is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Journal of Educational Supervision by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.

Principals' Perceptions of Teacher Evaluation Reform from Structural and Human Resource Perspectives

Journal of Educational Supervision

58 – 77

Volume 2, Issue 1, 2019

DOI: <https://doi.org/10.31045/jes.2.1.4>

<https://digitalcommons.library.umaine.edu/jes/>

John Wilson Campbell¹ and Mary Lynne Derrington²

Abstract

Driven by Race to the Top funding and quickly designed and deployed in 2010-2011, a new teacher evaluation policy in Tennessee altered principals' supervisory practices regarding their use of time for observation and reporting, their interaction with teachers, and the methods for giving teachers performance ratings. In addition, student test score data were integrated into final ratings, and professional consequences were linked with those ratings. Researchers in this study followed fourteen school principals over a five-year period to understand how their perceptions of new evaluation policy components affected their implementation. Data were analyzed using the structural and human resource frames of Bolman and Deal (2017) to interpret policy demands, principals' perceptions, and variations in implementation. Findings indicate that principals appeared to reject strict application of the policy's structurally focused components and procedures, while supporting collaborative, human resource-oriented approaches for promoting teacher growth and professional development.

Keywords

supervision; teacher evaluation; organizational theory; staff development; principals

¹ Alcoa City School District

² University of Tennessee

Corresponding Author:

John Wilson Campbell (Curriculum and Instruction, Alcoa City School District, 524 Faraday St., Alcoa, TN 37701, USA)

Email: jcampbell@alcoaschools.net

Introduction

When a half-billion-dollar federal Race to the Top (RTTT) grant led to Tennessee's teacher evaluation system overhaul in 2011, conflict arose across the state. Initial cooperation among policymakers, state and school district leaders, and teacher representation groups gave way to differences of opinion on the design, intents, and expectations of the new policy. School principals – immediately tasked with implementation and held accountable for the fidelity of that implementation – began the work of integrating challenging policy mandates into their supervisory practices. In 2011, this longitudinal study began to follow a group of principals over a five-year period as they struggled with the extensive new requirements and high-stakes consequences of the policy.

The new policies required principals to implement rigorous observation and scoring procedures, which would feed into a state-level teacher effectiveness calculation system. Given the central role of teacher evaluation in principals' supervisory responsibilities (and the way that role became embedded in the context of emerging conflicts over the new policy) the researchers began to question how principals were balancing their own local supervisory beliefs and approaches with the new state-level requirements and systematized teacher rating mechanisms. As a result, the researchers designed a longitudinal study to probe principals' perceptions of aspects of the new policy that appeared beneficial or problematic to them and to look at the way those perceptions affected their implementation of the new policy requirements. This study utilized the structural and human resource components of Bolman and Deal's (2017) multi-perspective organizational research as an analytical framework.

The purpose of this five-year, longitudinal, qualitative study was to examine how principals' perceptions of new teacher evaluation policy affected their implementation. By applying two different perspectives, the structural and human resource frames of Bolman and Deal (2017), the study provides a profile of principals' evolving perceptions and practices when implementing a new, rigorous state policy. The study asked the following research question:

How did principals' perceptions of a new teacher evaluation policy's components and expectations affect their implementation of the policy?

Literature Review

The United States educational reform environment, particularly during the past three decades, led to new education reform paradigms which call attention to differing organizational approaches. Performance expectations and pressures – enforced through test-based accountability programs and competitive, market-based principles applied to public education – have assumed a powerful role in efforts to improve educational outcomes (Mathis & Trujillo, 2016; Scott & Jabbar, 2014). This has been driven by large-scale federal accountability programs like No Child Left Behind (NCLB) and federal grant and accountability programs such as RTTT.

As these new policy expectations – drawn up by policymakers at some distance from the schools – play out in myriad complex and variable school settings, a degree of local interpretation and adaptive execution would be expected to influence implementation and the final effects (Honig,

2006). Looking specifically at local effects of RTTT-driven teacher evaluation reform in Tennessee, this study explores reform policy impacts through the eyes of principals whose work supervising teachers occurs at the juncture of state-imposed policy with the internal norms and culture of their organizations. This study's organizational behavior lens (Bolman & Deal, 2017) brings the RTTT-driven reform climate, teacher evaluation reform in Tennessee, the literature on supervision best practices, and the actual experiences of principals leading the implementation to a focal point.

Tennessee's Teacher Evaluation Overhaul and Initial Response

The RTTT-inspired, wholly redesigned teacher evaluation framework in Tennessee (Tennessee Department of Education, 2019; Tennessee State Board of Education, 2019) – encoded into state law, with accompanying high-stakes consequences – placed relatively heavy emphasis on the technical aspects and quantitative outcomes of frequent supervision and evaluation cycles. Designed in response to the overwhelming evidence that teacher evaluation nationwide was ineffective (Danielson & McGreal, 2000; Darling-Hammond, 2013; Kennedy, 2010; Marzano, Frontier, & Livingston, 2011; Popham, 1988; Weisberg, Sexton, Mulhern, & Keeling, 2009), the new policies required a) multiple annual observations of every teacher; b) significant scoring, feedback, and reporting requirements; c) new accountability linkages with student test score results; and d) potential professional consequences (e.g., gain or loss of tenure, retention, bonus pay).

The new policy required principals to use a new, complex, 4-domain, 23-part indicator instructional rubric (based at the time on NIET's TAP rubric) to score teachers and to report the observation ratings directly to the state. Pacing guides specified which portions of the rubric would be used and when, how many observations of each domain would occur and when, the total number of observations, and the minimum requirements for the total amount of observation time in classrooms. State training called for principals to score the indicators objectively (i.e., score each indicator a 1, 2, 3, 4, or 5 based on the degree to which events observed at the time of observation met the standards described in the rubric). Every observation during the year generated scores. At the end of the year, all scores were automatically averaged for all indicators to generate a summative, "qualitative," observation score.

This requirement created a more distinct management hierarchy between principals and teachers. Previously, a principal would conduct and record teacher performance appraisals once every five years and share results with only the teacher being evaluated. The principal filed the record at the school and was the final authority on the teacher's rating. There was no requirement to use or consider student test score data in the appraisal, no formula for including it in any rating, and no formal linkage between evaluation outcomes and either quantitative performance outcomes or professional benefits (e.g., tenure, retention).

The new policy's observation and scoring system structurally repositioned principals from the school's evaluation authority role to a mid-level scoring and reporting role. Now, principals would conduct every teacher's multiple, annual scored observations and report results directly to the state through a digital interface to be mathematically merged (weighted at 50%) with value-added testing and other achievement data (weighted at 50%) by a third-party software vendor for

calculation of final effectiveness scores. Proportional weighting exceptions were made over time due to state testing irregularities and the inability to apply the exact same set of measures to every type of teaching assignment. As a part of the evaluation system overhaul, state law linked professional consequences such as tenure and teacher dismissal to teachers' annual effectiveness ratings. Previously, principals held nearly full authority for these actions. In addition, bonus pay systems and differentiated pay scales, advocated for by state leaders at the time and required of all school districts, were linked with the new effectiveness rating system.

As the new policy went into effect in the fall of 2011, teachers, administrators, and other stakeholders responded with growing discomfort, launching a range of questions, concerns, and frustrations, which were often expressed publicly (Heitin, 2011; Tennessee Department of Education, 2012). Local media followed concerns expressed in school board meetings and other public forums (Garland, 2012; Johnson, 2014). Specific concerns included the complexity of the rubric, ambiguities in the rubric, teachers' planning time requirements, principals' time constraints, the state's rushed implementation, incomplete components, training, support systems, flaws in value-added scoring systems, and some teachers' appraisals being based on other teachers' test performance.

School principals, forced to realign their supervisory responsibilities, shouldered the burden of putting the new policy into action with their teachers and managing the impact of its operation. The policy's systemic complexity, its principal-as-state-observer-and-rater role, and its emerging political challenges tested principals' abilities to adjust their supervisory approaches and execute the state policy while preserving valued, locally developed leadership qualities previously established as a part of their schools' professional climate. The literature on supervision and teacher evaluation describes principals' supervisory responsibilities and best practices. Considering this scholarship is a key to understanding how the new policy affected principals' perceptions and implementation.

Complexity of Supervision

Principals' supervision of teachers occurs in a complex, multi-faceted, socially situated environment. Lortie (2009), describing the difficulties of evaluation, said, "school management is fundamentally interactive in nature," and that "the major complexities that emerge focus on relationships with other people" (p. 124). Supervision, of which teacher evaluation is a part, is now better understood as a complex combination of applied skills, which include developing curriculum, leading professional development, facilitating change, and addressing diversity (Glickman, Gordon, & Ross-Gordon, 2013). Furthermore, specific supervision and evaluation tasks might be mandated, yet effective supervision "must be flexible, balanced, adaptable, and shaped by the people who apply it" (Zepeda, 2017, p. 6). In addition, the school principal, who is ultimately responsible for teacher development, is responsible for myriad other academic and non-academic obligations. Consequently, teacher evaluation is best understood holistically "as part of a teaching and learning system" (Darling-Hammond, 2013, p. 3). This work includes differentiated teacher support, instructional feedback, teacher development, and cultivating collaborative, supportive adult learning environments (Glickman et al., 2013). Zepeda (2017) situates the basic tasks of clinical supervision within the richer context of interactive, empowered, collaborative learning cultures that truly support improved learning outcomes.

In contrast, supervision has also been viewed as an “instrument for controlling teachers” (Glickman et al., 2013, p. 8) in part through its link with evaluation. Zepeda (2017) discussed weaknesses of clinical supervision models in practice, which resulted in ineffective routines or hierarchical control. High-stakes consequences attached to evaluation routines tend “to shut down adult learning,” as “evaluation instruments often get in the way” (Marshall, 2006, p. 730). Education reform paradigms of the past decade may to some degree mirror these contrasts. School principals play a central role in making sense of the contrasting demands of internal operations and culture and externally imposed reform policies (Honig & Hatch, 2004).

Variation from Policy Designs

This study assumes that policy implementation would vary from policy design, as implementation literature describes (Berman & McLaughlin, 1978; Datnow, 2006; Honig & Hatch, 2004). Beyond simply chronicling an example of such variation, Bolman and Deal’s (2017) frames approach was used to analyze the *nature* of principals’ policy adaptation – the perceptions they developed about the policy and the ways those perceptions affected their implementation of the policy. Considering the structure and reach of the new evaluation policy against the backdrop of the complex supervisory environments into which it was introduced, implementation and fidelity challenges should be predicted.

It is worth questioning the degree to which the new evaluation policies aligned with concepts of supervision described above, for example, by Darling-Hammond (2013), Glickman et al. (2013), and Zepeda (2017). Zepeda (2017) said, “healthy school cultures thrive in environments built through collaboration, trust, and care for members of the school” (p. 8). Glickman et al. (2013) described effective supervision as, in part, “a collegial rather than hierarchical relationship between teachers and formally designated supervisors” (p. 7). Thus, effective school leaders could be expected to alter policy demands if such demands obstructed their best supervisory judgments. Applied specifically to the teacher evaluation policy examined in this study, principals experiencing both congruence and conflict with state policy prescriptions and local, relational priorities would come as no surprise. Resolving the different factors within such conflict is aided by adopting contrasting points of view. Bolman and Deal’s (2017) organizational frames provide this multiple perspective approach by reflecting both the forces of the policy and the forces of principals’ supervisory judgment.

Organizational Behavior: Structural and Human Resource Perspectives

To guide the investigation of principals’ policy perspectives and implementation, in light of the complex supervisory environment in which principals work, the researchers chose to utilize the organizational theory frameworks of Bolman and Deal (2017). Bolman and Deal developed an extensive body of work examining organizational events, structures, behavior, and decision-making from four different perspectives (structural, human resource, political, and symbolic). They built their work on organizational theory, sociology, psychology, and other fields, reflected in their development of four perspectives, or “frames.” Powerful, effective management strategies are embedded within each frame, with none representing a better or worse approach. It could be argued that all four Bolman and Deal (2017) frames must be utilized in concert for a

complete analysis. For the purposes of this article, however, the researchers propose an analytical starting point, basing the analysis on the two most salient frames from principals' experiences – the structural and the human resource frames – to lay a foundation for further study.

According to Bolman and Deal (2017), structural forces tend to reinforce hierarchical coordination, objective and rational tools and procedures, and the goal-oriented work of organizational structures. The structural position originated in the development of “scientific management” techniques (Taylor, 1911, cited by Bolman & Deal, 2008), which “broke tasks into minute parts and retrained workers to get the most from each motion” (p. 48). The human resource perspective, on the other hand, reinforces the needs of individuals within the organization and places a high value on individual talents and ideas. Human resource approaches “invest in learning” and “encourage autonomy and participation” (Bolman & Deal, 2008, p. 142). Bolman and Deal assert that effective organizational work requires the right blend of forces described by these frames.

An examination of the structural and human resource aspects of teacher evaluation policy implementation begins with reference to the Professional Standards for Educational Leaders (National Policy Board for Educational Administration [NPBEA], 2015). Tennessee principals are trained and certified to follow its guidelines. The standards delineate both structural and human resource-oriented imperatives.

Supervision: Structural Perspectives

The NPBEA Professional Standards for Educational Leaders convey the importance of structurally sound school oversight, with phrases such as, “institute, manage, and monitor operations,” “strategically manage... resources,” “know, comply with... local, state, and federal laws...,” “develop and manage,” and “manage governance processes.” (National Policy Board for Educational Administration, 2015, p. 17). Effective leadership does require an effectively managed, intentional, accountability-driven approach. Glickman et al. (2013) used the term “technical supervisory skills” (p. 145) for ways effective leaders structure their work.

Glickman et al. (2013) also described, however, problems “aggravated by the external control brought about by legislated reform” (p. 25). In addition to the structural complexities of the state evaluation policy's rubric design, time requirements, observation, scoring protocols, reporting, and mathematical formulae for merging test and observation data, professional incentives and disincentives (e.g., tenure, bonus pay) were tied directly to the calculated scores and rankings of teachers' overall effectiveness. A number of researchers have questioned the validity and effectiveness of using student test data for teacher ratings (Berliner & Glass, 2014; Popham, 2013; Popham & DeSander, 2014). Scoring processes that would affect pay, promotion, retention, or tenure have been affected by reliability and validity concerns (Berliner & Glass, 2014; Holloway-Libell, Amrein-Beardsley, & Collins, 2012). Linking pay to evaluation ratings, particularly with ratings tied to students' value-added scores, has been viewed as an untrustworthy policy component and perhaps as less effective for teacher improvement than initially intended (Goldring et al., 2015). Looking across multiple evaluation models, Steinberg and Kraft (2017) found significant variations in the teacher performance cut scores and the

weighting of evaluation components embedded in various policy designs, implying that structural consequences for a teacher might stem more from policy mechanisms than instructional actions.

Darling-Hammond (2013) expressed concern for the impact of some quantitative rating policies on the development of teacher collaboration, noting that using competitive principles in evaluations systems (e.g., scoring, ranking) can “undermine the growth of learning communities” (p. 3) and that “efforts to evaluate and compensate teachers based directly on students’ test scores can create unintended, dysfunctional consequences” (p. 62). The rewards of high test and observation scores, or the negative consequences of lower scores can become comingled with observation scoring requirements, affecting principals’ responses to implementation.

Supervision: Human Resource Perspectives

With regard to human resource aspects of supervision, NPBEA standards (National Policy Board for Educational Administration, 2015) include, “Empower and entrust teachers and staff with collective responsibility,” “Establish and sustain a professional culture of engagement and commitment to shared vision, goals, and objectives,” “Develop and support open, productive, caring, and trusting working relationships among leaders, faculty, and staff to promote professional capacity and the improvement of practice,” and “collegial feedback, and collective learning” (p. 15). The introductory material acknowledges “the central importance of human relationships not only in leadership work, but in teaching and student learning” (p. 3). Such standards are tightly aligned with supervision research that suggests principals should implement teacher evaluation to serve as supportive, collaborative, instructional skill builders and engaged leaders of professional development (Glickman et al., 2013; Louis & Robinson, 2012; Spillane, Diamond, & Burch, 2002; Stein & Nelson, 2003; Zepeda, 2017).

Zepeda (2017) discussed the importance of effective supervisors creating “the conditions that support motivated teachers,” demonstrating a “sense of empathy,” and working to “individualize learning opportunities” (p. 148). Expanding on instructional leadership, Stein and Nelson (2003) wrote of the importance of principals’ knowledge of creating connections between instruction, subject matter content, and the learning process. Research from multiple sources describes effective school leadership in terms that focus on intense, understanding, supportive, and improvement-driven relationships between school leaders and their teachers (Glickman et al., 2013; Honig, 2012; Spillane, Halverson, & Diamond, 2004). The teaching and learning process in schools is, by its very nature, complex and highly contextualized, with improvement efforts often based on particular circumstances and non-routine events (Glickman et al., 2013; Supovitz, 2006).

Multiple Imperatives

Supervision literature provides guidance for improving schools from both structural and human resource approaches by promoting well-designed, structural approaches for cultivating rich, relational cultures that drive both adult and student learning (Darling-Hammond, 2013; Glickman et al., 2013; Zepeda, 2017). Just as qualities of Bolman and Deal’s (2017) structural and human resource frames must function together for organizational effectiveness, so too

should structural, organizing imperatives undergird policies and practices that focus on the human resource development required for effective supervision.

Kennedy (2010) noted that “we hold numerous public values, or criteria for defining teacher quality, and sometimes they contradict one another” (p. 226). Cohen, Spillane, and Peurach (2017) described fundamental conflicts, “because school systems are open systems and depend on their environments for students, funds, political support, guidance, and legitimacy... yet they operate in a complex, pluralistic institutional environment that contains different and often divergent pressures for action” (p. 209). Newly legislated, structurally focused, formula-driven teacher quality reform policies prevailed in 2011, and principals assumed central mediating positions with their faculties, balancing mandated, externally produced procedural expectations and formulae with more personalized strategies for successful implementation in their schools (Flores & Derrington, 2017).

Research Design and Methodology

A longitudinal, exploratory, qualitative methodology was used to collect data from principals of 14 schools as they interpreted and implemented new teacher evaluation policies. The methodology was chosen to best understand the way the participants created meaning based on their perceptions and interpretation of experiences over time (Derrington, 2018; Merriam, 2009). Additionally, returning multiple times to the same participants enriched the data and improved the credibility of analysis over time by revisiting participants’ previous thoughts, looking for changes in perceptions and actions, seeking “saturation” points in the data (when certain themes became repetitious [Merriam, 2009]), and observing subtle changes in these themes over time.

Participants and Settings

The sample was purposeful to serve the research objectives and to provide the most information (Patton, 2002). The selection criteria included three or more years as a principal so that the evaluation of implementation would not be confused with challenges common to new principals. The selection of principals represented rural and suburban contexts, similar to a preponderance of school districts in this east region of Tennessee. Urban, large city districts – of the type that drive many implementation studies – were not represented. The schools were also chosen to represent various grade-level configurations in order to examine possible heterogeneity (Patton, 2002). Depending on district size, one or two high schools, middle schools, and elementary schools were chosen in each district so that similarities and differences might be considered by grade level. Additionally, the schools represented the widest possible variation in student demographics among the sample districts. Free and reduced lunch percentages ranged from 31% to 85% among the selected schools. Student population varied from the smallest school with 295 students to the largest with 1,486 students. Table 1 (below) details the schools’ demographics. Fourteen schools provided an adequate research sample and included four high schools, five middle schools, and five elementary schools. Eight of the participants were male; six were female. Informed by the research questions, interview protocols were constructed based on current literature.

Table 1. Demographic Data of Participating Schools

District	School	Enrollment	Grades Served	Free and Reduced Lunch Percentage
District A	Elementary	513	PK-4	57%
	Middle	735	7-8	31%
	High	1,486	9-12	26%
District B	Elementary	614	PK-5	85%
	Middle	362	4-8	75%
	High	1,245	9-12	48%
District C	Elementary	568	PK-5	41%
	Middle	326	PK & 6-8	71%
	High	1,479	9-12	53%
District D	Elementary 1	691	PK-4	51%
	Elementary 2	410	PK-5	85%
	Middle 1	822	5-8	44%
	Middle 2	295	6-8	81%
	High	711	9-12	67%

Note. Participating school data is based on the 2011 Tennessee Department of Education Report Card and school websites. Decimal numbers have been rounded.

Data Coding and Analysis

Principals were interviewed from 50 to over 60 minutes each year of the study. The interviews were primarily face-to-face. Fifty-six total hours of interviews were recorded with consent and were transcribed verbatim.

Transcripts were coded manually using abbreviations to label segments of participants' responses. Manual coding can be tedious (Miles & Huberman, 1994), but was chosen for ease of multiple re-readings and discussion between the researchers. Regardless of the coding technique, Miles and Huberman emphasized the importance of structural order, which was consistently maintained throughout the five years. Interview data were coded and grouped or "chunked" (Creswell, 2003; Johnson & Christensen, 2004) to begin identifying and developing themes. A master list was applied when the data segments fit the codes. New codes were generated when data segments did not fit existing codes. Codes included abbreviations for participants' perceived concerns and benefits during implementation. Many coded themes were identified during the early stages of the study, during initial interviews and the subsequent reading and rereading of

transcripts. As the study progressed over five years, early analyses were revisited, the evolution of initial themes were identified, and new themes were developed, based on the appearance of new topics in the data. Using the same method of data collection and analysis, applied repeatedly over time, allowed the researchers to analyze how many issues had changed (Flick, 2014).

Findings

The researchers utilized the organizational behavior frames of Bolman and Deal (2017) to answer the research question, how principals' perceptions of a new teacher evaluation policy's components and expectations affect their implementation of the policy. Data were generated from interview questions designed to understand principals' perception of policy requirements and evaluation expectations. Two of Bolman and Deal's organizational behavior frames, the structural frame and the human resource frame, were used to analyze the findings.

Over five years of data collection, starting with the first year of implementation, the participating principals shared a wide range of experiences, reactions, and practices. From the collected data, the researchers developed four salient themes that represent the predominant perceptions and associated implementation practices across multiple participants. These four themes are: Time management and policy requirements, application of the rubric and observation protocols, state-calculated teacher ratings, and structured and unintended consequences.

Time Management and Policy Requirements

The amount of time required to complete procedural expectations challenged principals from a structural perspective, pushing their existing human resource priorities to the periphery as policy implementation began. From the structural perspective, the immediate escalation of time required to implement the policy's expectations was a top concern in the first years of implementation. The previous requirement to conduct two low-stakes evaluations every ten years for each teacher changed to multiple observations with detailed feedback (as part of a full, scored, annual evaluation) for every teacher. The situation was aggravated by the policy's rapid design, minimal piloting, and the immediate expectation of policy compliance. Principals were still learning about all the requirements while being held accountable for implementation. Principal comments included, "You've got to find the hours somewhere, and it's hard to do. Just the volume of paperwork itself... I've spent a lot of time before school, after school, and weekends," and "It has taken a lot of time outside of working hours; I mean weekends, nights, breaks." Another said, "I've not done a good job of balancing all of my responsibilities... it's been evaluations all the time... you put your head down and you went at it."

Over time, principals described routines developed to manage the expectations, including learning the rubric and working with assistant principals to distribute the work load and set schedules for the year. Participant comments included, "...principals realize the time commitment, and they schedule that out," "I think we just learned how to stay on top of it," and, "you about have to have the rubric memorized when you start doing scoring, and that takes less time."

As the study progressed over five years, concern about time demands eased and reflected a shift from structural logistics concerns toward the human resource perspective. Principals increasingly noted the benefits of multiple observations, despite the time requirements, which remained, but became more routine. By year five, most principals reported that even though the demand on time remained very high, the process “was not new,” or had become a “routine,” which indicated integration of the policy’s structural requirements into their work day.

Human resource benefits began appearing in the data and became more common by year five. One principal said, “I have had an opportunity to have lots of really good conversations with teachers that might not have happened otherwise.” Another said, “I’m looking a little deeper at what I’m seeing in the classroom, which makes my comments a little deeper, and that takes a little longer to formulate.” A third said, “You’ve got to invest the time to get the results.” Multiple principals suggested, however, being unable to continue making frequent, informal classroom visits, which they valued. The sentiment was captured by one principal’s comment that “the process has actually taken us out of the classroom because previously we were there every day... that is a complaint from teachers.”

For principals already working a complex job, the impact of heavy, new structural demands challenged their management skills and their previously established staff relations perspectives, with regard to being present among their teachers – in the classrooms and hallways. Over time, however, they worked to simplify the structural requirements (e.g., memorize the rubric, build schedules, delegate responsibilities) and appeared to weave their more relational, human resource-oriented values (e.g., conversation, investing time in feedback) into the structural demands of the policy.

Application of the Rubric and Observation Protocols

Teacher observations were to be conducted using the 23-item rubric. Specified indicators were to be scored as objectively as possible within each observation, using only the observation itself and, possibly, pre-observation conference documentation. Structurally, though it took a year or two to learn, principals quickly supported the contents of the rubric. Eighty percent of participants cited the rubric as a benefit of the new policy: a level of support that was sustained throughout the study. Principal comments included, “...it specifically spells out what you have to do, and I relate that to teachers,” “it’s just gotten teachers and administrators both... recognizing good instruction,” and, “focusing on teacher evaluation as well as those best practices in the classroom has actually increased and continues to increase teacher awareness.” Early in the study, principals reported teachers and administrators using the rubric as a focus for collaborative staff development. Late in the study, a principal remarked, “we have something that can take us a long way.” The rubric appeared to offer a structured index of multiple, desirable classroom qualities, packaging them in a readily accessible format, with common terms and expectations for principals to use as instructional leaders and human resource developers.

Over time, even though support for the structured collection of observable qualities remained high, principals expressed dubious acceptance of the required observation and rating method. Exerting a human resource-oriented perspective, they reported making personal decisions about rating levels or whether to record what they observed on any particular classroom visit.

Principals appeared to lack confidence in interpretation or hesitance to assign low ratings when there were reasons to believe they were not deserved or appropriate. The observation and rating regime grew to become a consistent concern.

Principal comments included, “I may rate some people higher than maybe they deserve because I’ve been working with them all this time,” and “I don’t ever want to beat anybody over the head with anything like this.” One principal commented:

We know our teachers. We know what is normal and abnormal with their procedures and teaching styles... if I walk into a classroom... and this is just really an off day for them... I’m going get up and leave and we’re going to do that some other time. That’s just the ethical thing to do, I think. We all have off days. We’re not setting them up for failure. We don’t want a gotcha game.

In the late years of the study, such instances of personal choice began to exemplify principals’ adoption of a human resource stance in opposition to a structural, policy-driven one. This resulted, however, in reports of inconsistency across schools. Principal comments included, “I know different systems still have different ideas of what teachers should score. Different schools use it as a strict guideline... and others use it more loosely,” and “it is subjective; there’s not much way to take out the variation between schools and between different evaluators.” One principal reported, “some will look at the rubric and take it very literally, whereas others kind of make assumptions or dig a little deeper through follow up conversation... affecting the scores that teachers are getting.” Another, describing a former teacher’s formerly high-rated instruction, who was now receiving lower scores at a new school, said, “nothing really has changed. That principal sees it completely differently.”

Over time the rubric and its application became the center of the tension between structural and human resource perspectives. Additionally, the instructionally valuable components and its use as a guide for instructional leadership were beneficial. However, the structural requirements for its use as a rating device appeared to disturb principals’ human resource values with regard to supervising their teachers.

State-Calculated Teacher Ratings

Structurally, the state’s teacher effectiveness rating system (using principals’ submitted observation ratings, state test-derived value-added scores, and other achievement measures) occurred at the state level, outside of the school, and at some distance from direct principal influence. The system generated a one-to-five “level of effectiveness” score for every teacher. (The policy included a structural expectation that teachers’ test-based value-added scores would align with their observation scores). From almost no mention in the first year of data collection to repeated mentions in later years, principals increasingly expressed concerns about accuracy and fairness to teachers, being sensitive to human resource qualities and their own perceived reduction of authority over final results.

Principal comments included, “one concern I have... is in the scoring... having test scores play a major role in a teacher’s evaluation score... we don’t always have control over that... and its one

snapshot of a student”, and, “talk about throwing darts in the dark... I see a teacher do a really good job... but on those four days where their students test, they don’t do so well.” Of value-added measures used for summative effectiveness calculations, one principal said of teachers, “their value-added score could have a million things that don’t have a whole lot to do with how much they [students] learned that year.” Some non-tested teachers were required to take school-wide test scores as the quantitative, test-based factor in their final effectiveness rating. This was seen as unreasonable. One principal said, “a lot of teachers might be a 5, but may not be doing level-5 work,” indicating that the structurally inflexible rating system misrepresents some teachers’ effectiveness.

Such data from the later years of the study indicated that principals viewed their supervisory responsibilities as the school’s human resource leader to have been interrupted by the state requirement. In place, principal’s view their work as a remotely managed, multi-factor, structural system. Lack of direct control over test scores and final ratings illustrated the shift of supervisory authority from the schools’ leaders to the state system.

Structured and Unintended Consequences

Related to concerns about using test scores in effectiveness calculations, principals expressed concerns about the policy’s consequences – both structured and unintended. With professional benefits such as bonus pay and tenure becoming a consequence of system-generated ratings calculations, they grew concerned about competition undermining collaboration. Concern for unintended outcomes grew, such as with the policy missing its intent in some cases due to its reliance on a structure of calculated ratings.

One principal commented, “my concern...[is with] what the state is asking us to consider doing with these results, for example on strategic compensation.” Another said:

What worries me is that we’re going to have our focus shifted if we start finding that your observation score has to fall in this range to get a 1% raise... or to get a step on the salary scale... it’s going to be more contentious than truly honest.

Also concerned about the focus on scores, but more with potentially undesirable effects on instruction, another said:

In the beginning, everybody was doing what they needed to do to get that five, and now they don’t see that as necessary. So they’re getting a bit more lax. Some of them try to figure things out numerically, and they’re seeing that an extra five here or there doesn’t impact as much as they thought earlier on.

These human resource concerns were not evident in the first years, as principals focused on managing the new requirements of the policy. As the study progressed, however, principals’ qualms with the consequences of these policy structures pointed to a declining level of perceived credibility. From the perspective of developing their teachers with a locally driven human resource perspective, policy mechanisms associated with consequences began to appear counterproductive or ineffective.

Findings Summary

Principals were consistently challenged by the policy's structurally heavy time demands, but valued time spent on human resource activity such as engaging in instructionally focused conversations. The rubric was supported as a common reference document that could be used in formative, professional learning activities. On the other hand, principals did not support using the rubric as a mandatory teacher-rating instrument. They reported exerting a prerogative to affect the scores they gave teachers by making procedural or interpretive decisions, which aligns with the human resources frame perspective. Principals did express doubt about consistent ratings from school to school but gave priority to their own right to apply the rubric in the manner that made the most sense to them in the context of their relationship with teachers.

Over time principals expressed a lack of confidence in the policy's structured test score rating components and calculation system. Confidence in the relationship between state test data and observed teacher performance appeared to be low. Principals were increasingly concerned about unintended effects of the quantitative, partially test-based, state-calculated ratings system that did not align with the human resource-oriented framework.

Discussion

The purpose of the study was to examine the ways principals' perceptions of new teacher evaluation policy affected implementation. The analysis applied the structural and human resource frames of Bolman and Deal (2017). This analytical approach provided two perspectives for considering principals' experiences as they learned and implemented a new, rigorous teacher evaluation policy designed and mandated by the state. Principals successfully adopted and valued some structural components (e.g., the rubric, observation time) but modified or questioned others (e.g., observation scoring protocols, state effectiveness calculations). Principals perceived that teachers were at risk of potentially negative, undeserved consequences due to structural components of the policy. They also perceived that their collegial relationships with teachers were at risk if they were to fully execute the state policy as designed. Thus, circumventing the structural requirements of the prescriptive system by modifying scoring protocols allowed them to exert more human resource-oriented supervisory values, which included maintaining collaborative relationships or differentiating the process by teacher need. This conclusion suggests that modifying the implementation of highly structured, state-driven, consequential evaluation policy is essential to principals for maintaining relationships within the supervisory principal-teacher dyad. It reinforces research suggesting that teacher evaluation is a complex, multi-faceted, and relationship driven supervisory activity (Glickman et al., 2013; Lortie, 2009; Zepeda, 2017).

Principals' qualms about observation and teacher scoring were related to the low levels of credibility ascribed to the teacher effectiveness calculations that would include those observation scores. In addition, doubts about the ability of student test results to reflect teacher quality contributed to overall concerns about outcomes for teachers. Critical of the validity of the final teacher rating process, it appeared that principals might have influenced that policy component through their handling of the observation scoring process. As indicated by this study, principals

protected teachers from punitive evaluation mandates by modifying or minimizing processes negatively linking ratings and test scores. This modification of structural mandates adds to previous research on the negative view of evaluation ratings linkage to value-added test score data as an untrustworthy policy component (Berliner & Glass, 2014; Holloway-Libell, Amrein-Beardsley, & Collins, 2012; Popham, 2013; Popham & DeSander, 2014), ineffective for teacher improvement (Goldring et al., 2015). This study began to describe how principals attempt to override the structural policy mechanisms with collegially supportive human resource approaches.

With the final summative ratings responsibility shifted from principal's supervisory responsibilities to a digital, state-run, mathematical calculation system, the policy intent of improving teacher performance is weakened. For instance, if a teacher's score would not trigger negative consequences, the score might be ignored, even if there were significant room for improvement. In the end, the policy's structurally dominant forces, driven by numbers and remote computer calculations, appeared to principals as only loosely related to their supervisory skills and efforts and to the actual performance of their teachers.

Conclusion

Utilizing Bolman and Deal's (2017) frames approach to study principals' efforts to implement a new, rigorous, state-mandated teacher evaluation policy clearly demonstrated that through their daily evaluation practices within this new policy environment, principals' human relations values outweighed their trust in the policy's structural components. The results of this analysis carry implications for policymakers, practitioners, and researchers. From a policymaking standpoint, it appears that structural attempts to reposition principals as teacher raters within a wider, high-stakes, state-run system, weakened or undermined their professional, supervisory authority. Persistent policy challenges appear to revolve around who determines a teacher's effectiveness and what factors are used to develop the judgment. The implication is that policymakers must consider the central role of principals as supervisors and evaluators prior to teacher evaluation redesign.

Furthermore, a key area for policy improvement is a consideration of the negative consequences and credibility of a system linking student test scores and high-stakes decisions such as tenure and merit pay. Credibility could be improved with a stronger human resource focused approach, guided local flexibility, and increased local authority for both teacher rating and development. Accountability might be better placed on supervisory excellence as opposed to procedural compliance. Policy designs that are more aligned with human resource focused leadership and with best practices from supervision literature might support and sustain more effective teacher evaluation systems.

From a practitioner standpoint, an implication is that a system of *guided* principal autonomy within the framework of a common, state-wide evaluation program is needed. Given the value found in the rubric and in the time committed to observation and feedback, it appears that school leaders need improved resources (e.g., rubrics, guides), training, and strong, locally developed accountability that recognizes and supports professional judgment and autonomy. Professional

learning for principals could focus less on structural, procedural compliance and more on human resource development and effective use of procedural flexibility.

Further research might explore to what extent any individual principal, or larger group of principals, varied implementation from specified policy, delving deeper into decision-making and its effect on teachers. Answering such questions may help determine the degree to which principal variation stemmed from constructive supervisory leadership or from principals shielding teachers from negative policy effects. If variation among principals is widespread, additional research focusing on student performance and teachers' summative ratings might help determine the degree to which teacher evaluation policy reform has positively affected instruction and outcomes.

This study examined evaluation implementation using the structural and human resource frames (Bolman & Deal, 2017). Bolman and Deal (2017) also developed political and symbolic frames of reference, and future studies might consider an examination of the findings from those perspectives as well. As Bolman and Deal note, certain aspects of any one frame can be over-developed in policy or procedure development creating imbalances, loss of desired impact, or organizational distrust or dysfunction. In the case of teacher evaluation reform in Tennessee, as documented in this study, considering multiple points of view from both state and local levels of leadership could help with developing and sustaining a credible, well-functioning evaluation policy, aligned vertically across state and local leaders.

References

- Berliner, D., & Glass, G. V. (2014). Chipping away reforms that don't make a difference. *Educational Leadership*, 71(9), 28-33. Retrieved from <http://www.ascd.org/publications/educational-leadership/summer14/vol71/num09/Chipping-Away@-Reforms-That-Don't-Make-a-Difference.aspx>
- Berman, P., & McLaughlin, M. W. (1978). *The RAND change agent study* (Vols. 1-8). Santa Monica, CA: The Rand Corporation.
- Bolman, L. G., & Deal, T. E. (2008). *Reframing organizations: Artistry, choice, and leadership* (4th ed.). San Francisco: Jossey-Bass.
- Bolman, L. G., & Deal, T. E. (2017). *Reframing organizations: Artistry, choice, and leadership* (6th ed.). San Francisco: Jossey-Bass.
- Cohen, D. K., Spillane, J. P., & Peurach, D. J. (2017). The dilemmas of educational reform. *Educational Researcher*, 47(3), 204-212. doi:10.3102/0013189X17743488
- Creswell, J. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Design.
- Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. New York: Teachers College Press.
- Datnow, A. (2006). Connections in the policy chain: The "co-construction" of implementation in Comprehensive School Reform. In M. Honig (Ed.), *New directions in education policy implementation*. (pp. 105-123). Albany, NY: State University of New York Press.
- Derrington, M. L. (2018). *Qualitative Longitudinal Methods: Researching Implementation and Change*. Thousand Oaks, CA: Sage.
- Flick, U. (2014). *An introduction to qualitative research*. Thousand Oaks, CA: Sage.
- Flores, M. A., & Derrington, M. L. (2017). School principals' views of teacher evaluation policy: lessons learned from two empirical studies. *International Journal of Leadership in Education*, 20(4), pp. 416-431. doi:10.1080/13603124.2015.1094144
- Garland, S. (2012, June 11). Widespread dissatisfaction with new teacher evaluations in Tennessee. *HechingerEd Blog*. Retrieved from http://hechingered.org/content/widespread-dissatisfaction-with-new-teacher-evaluations-in-tennessee_5126/
- Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2013). *The basic guide to supervision and instructional leadership* (3rd ed.). Upper Saddle River, NJ: Pearson.
- Goldring, E., Grissom, J.A., Rubin, M., Neumerski, C.M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44, 96-104. doi:10.3102/0013189X15575031
- Heitin, L. (2011, October 18). Teacher-evaluation rush may jinx other states' efforts. *Education Week*. Retrieved from <https://www.edweek.org/ew/articles/2011/10/19/08eval.h31.html>
- Holloway-Libell, J., Amrein-Beardsley, A., & Collins, C. (2012) All hat and no cattle. *Educational Leadership*, 70(3), 65-68. Retrieved from <http://www.ascd.org/publications/educational-leadership/nov12/vol70/num03/All-Hat-and-No-Cattle.aspx>
- Honig, M. I. (2006). Complexity and policy implementation: Challenges and opportunities for the field. In M. Honig (Ed.), *New directions in education policy implementation*. (pp. 1-23). Albany, NY: State University of New York Press.

- Honig, M. I. (2012). District Central Office Leadership as Teaching: How Central Office Administrators Support Principals' Development as Instructional Leaders. *Educational Administration Quarterly*, 48(4), 733-774. doi:10.1177/0013161X12443258
- Honig M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple external demands. *Educational Researcher*, 33(8), 16-30. doi:10.3102/0013189X033008016
- Johnson, L. (2014). Tennessee teachers push back on evaluation process. *Huffington Post*. Retrieved from http://www.huffingtonpost.com/2014/02/10/tennessee-teachers-evaluation_n_4760096.html
- Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches*. Boston, MA: Pearson Education.
- Kennedy, M. (2010). *Teacher assessment and the quest for teacher quality*. San Francisco, CA: Jossey-Bass.
- Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 46(5), 234-249. doi:10.3102/0013189X17718797
- Lortie, D. C. (2009). *School principal: Managing in public*. Chicago: University of Chicago Press.
- Louis, K. S., & Robinson, V. M. (2012). External mandates and instructional leadership: School leaders as mediating agents. *Journal of Educational Administration*, 50(5), 629-665. doi:10.1108/09578231211249853
- Marshall, K. (2006). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan*, 86(10), 727-735. doi:10.1177/003172170508601004
- Marshall, K. (2009). *Rethinking teacher supervision and evaluation*. San Francisco, CA: Jossey-Bass.
- Marzano, R., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Design.
- Mathis, W. J., & Trujillo, T. M. (2016). *Learning from federal market-based reforms*. Charlotte, NC: Information Age Publishing.
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.
- National Policy Board for Educational Administration (2015). *Professional standards for educational leaders*. Retrieved from <https://www.npbea.org>
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Popham, W. J. (1988). The dysfunctional marriage of formative and summative teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(3), 269-273. Retrieved from <https://link-springer-com.proxy.lib.utk.edu/journal/11092>
- Popham, W. J. (2013). Can classroom assessments of student growth be credibly used to evaluate teachers? *English Journal*, 103(1), 34-39. Retrieved from <http://www.ncte.org/>
- Popham, W. J., & DeSander, M. (2014). Will the courts save teachers? *Educational Leadership*, 71(5), 55-58. Retrieved from <http://www.ascd.org/publications/educational-leadership/feb14/vol71/num05/Will-the-Courts-Save-Teachers%20A2.aspx>

- Rutledge, S. A., Harris, D. A., & Ingle, W. K. (2010). How principals “bridge and buffer” the new demands of teacher quality accountability: A mixed methods analysis of teacher hiring. *American Journal of Education*, 116, 211-242. doi:10.1086/649492
- Scott, J., & Jabbar, H. (2014). The hub and the spokes: Foundations, intermediary organizations, incentivist reforms, and the politics of research evidence. *Educational Policy*, 28(2), 233-257. doi:10.1177/0895904813515327
- Spillane, J. P., Diamond, J. B., & Burch, P. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, 16(5), 731-762. doi:10.1177/089590402237311
- Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies*, 36(1), 3-34. doi:10.1080/0022027032000106726
- Stein, M. K., & Nelson, B. S. (2003). Leadership content knowledge. *Educational Evaluation and Policy Analysis*, 25(4), 423-448. doi:10.3102/01623737025004423
- Steinberg, M. P., & Kraft, M. A. (2017). The sensitivity of teacher performance ratings to the design of teacher evaluation systems. *Educational Researcher*, 46(7), 378-396. doi:10.3102/0013189X17726752
- Supovitz, J. A. (2006). *The case for district-based reform*. Cambridge, MA: Harvard Education Press.
- Taylor, F. (1911). *The principals of scientific management*. New York: Harper.
- Tennessee Department of Education. (2012). Teacher evaluation in Tennessee: A report on year 1 implementation. Retrieved from <https://www2.ed.gov/programs/racetothetop/communities/tle2-year-1-evaluation-report.pdf>
- Tennessee Department of Education. (2019). Tennessee Educator Acceleration Model. Retrieved from <https://team-tn.org/evaluation/teacher-evaluation/>
- Tennessee State Board of Education. (2019). Teacher and Administrator Evaluation Policy. Retrieved from <https://www.tn.gov/content/dam/tn/stateboardofeducation/documents/pendingrulespolicies/october-19%2c2018%2f5.201%20Teacher%20Administrator%20Evaluation%20Policy%2010-19-18.pdf>
- Weisburg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. Retrieved from https://tnnp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf
- Zepeda, S. J. (2017). *Instructional supervision: Applying tools and concepts* (4th ed.). New York: Routledge.

Author Biographies

John W. Campbell is Director of Curriculum and Instruction for Alcoa City Schools, Alcoa, Tennessee and is President of the AIMS/TIGER teacher evaluation consortium in Tennessee, one of the state’s alternative teacher evaluation models. His research interests include teacher evaluation policy, politics in education, and organizational behavior. He has co-authored many refereed articles and presented papers at multiple national conferences.

Mary Lynne Derrington is an Associate Professor in the department of Educational Leadership and Policy Studies at the University of Tennessee. Her research interests include teacher evaluation, policy implementation, and global school reform issues. She has published numerous

refereed articles, presented papers internationally, and been an invited speaker on leadership and teacher evaluation.